

Office of the Chief Scientist for Human Factors

Vertical Flight Human Factors

Program Review
FY05



Dr. William K. Krebs, Aviation Maintenance Human Factors Program Manager

**Federal Aviation Administration
ATO-P R&D HF (Room 907A)
800 Independence Avenue, S.W.
Washington, D.C. 20591
phone (202) 267-8758
e-mail: william.krebs@faa.gov
<http://www.hf.faa.gov/krebs>**

The Federal Aviation Administration Office of the Chief Scientific and Technical Advisor for Human Factors (ATO-P R&D HF) vertical flight human factors program is a relative new research domain. Research in this area is meant to identify specific human factors associated with helicopter flight regimes within the National Airspace System. Such issues include certification and regulation of civilian flights with night-vision-goggles devices, simultaneous non-interfering operations, and implications of tilt-rotor controls.

The following report summarizes projects between October 1st, 2004 and September 30th, 2005. These projects attempt to address requirements identified by the Federal Aviation Administration Flight Standards and Certification offices. The intent of this report is to allow Federal Aviation Administration sponsors to determine whether their requirements have been satisfactorily addressed, allow investigators to receive feedback from Federal Aviation Administration sponsors and other interested parties, and to provide feedback to the ATO-P R&D HF vertical flight human factors program manager on the quality of the research program. Basically, this document is a means of holding each group (sponsor, investigator, ATO-P R&D HF program manager) accountable to ensure that the program is successful.

The FY05 funded projects had \$250,000 contract dollars distributed to three projects.

Additional information about the Vertical Flight Human Factors research program can be found at <http://www.hf.faa.gov/vertical.htm>

William K. Krebs, Ph.D.

<http://www.hf.faa.gov/krebs>

**Project Justification:
Vertical Flight Requirements Mapped to Projects**

The table below lists each project with the corresponding research requirement. Please click on the research requirement link to understand the FAA sponsor research need for the project

Project Title	Requirement Statement	Sponsor	Research Requirement link
Investigation ATC Procedures for Simultaneous Non-Interfering Flight Within the National Airspace System	To determine NAV performance of VFR helo pilots using IFR qualified GPS receivers. AFS needs to quantify helo pilot NAV performance for IFR and VFR pilotage which will allow the development of procedures to integrate within the national airspace system.	AFS-800	link
UAV See and Avoid Systems: Modeling Human Visual Detection and Identification	This research will compile and review the characteristics and performance of existing optical systems that could be used to enhance the human UAV operator's ability to see-and-avoid potential conflicts with other manned and unmanned aircraft. Data will be collected for those sensor systems that are currently being used in Commercial UAV operations (e.g., surveillance, search-and-rescue, law enforcement, etc.) to determine their ability to be used to detect and avoid conflicting aircraft. The types of systems (cameras) will be characterized by their performance characteristics: field-of-view, field-of-regard, modulation transfer function, focal point, and lens quality. This comparison will be used to determine the ability of these systems to detect static images of differing sizes, at a range of distances in, variety of visibility conditions, i.e., sense-and-avoid. Existing optical models will be used to analyze the performance of these systems for detecting when the optics are integrated with a processor and data link system to determine the effects of bandwidth, image compression, and latency on see-and-avoid performance for large and small conflicting aircraft operating at a range of speeds with both vertical and horizontal path variations leading to the conflict. Finally, the utilization of these systems will be evaluated	AFS-800	link

	considering the performance of the human operator's eyes in the role of see-and-avoid (human-in-the-loop).		
Video Processing Methods for In-Flight Gaze Analysis	To determine NAV performance of VFR helo pilots using IFR qualified GPS receivers. AFS needs to quantify helo pilot NAV performance for IFR and VFR pilotage which will allow the development of procedures to integrate within the national airspace system.	AFS-800	link

Vertical Flight Human Factors

FY05 Funded Projects

<u>Project Title</u>	<u>Page #</u>
Investigating ATC Procedures for Simultaneous Non-Interfering Flight Within the National Airspace System <i>Sullivan, J. & Darken, R.</i> <i>Naval Postgraduate School</i>	6
UAV See and Avoid Systems: Modeling Human Visual Detection and Identification <i>Watson, A. B.</i> <i>NASA Ames Research Center</i>	11
Video Processing Methods for In-Flight Gaze Analysis <i>Mulligan, J. B.</i> <i>NASA Ames Research Center</i>	18

INVESTIGATING ATC PROCEDURES FOR SIMULTANEOUS NON-INTERFERING FLIGHT WITHIN THE NATIONAL AIRSPACE SYSTEM

Principal Investigators:
CDR J. Sullivan USN & Dr. R. Darken,
700 Dyer Rd. Code MOVES
Naval Postgraduate School
Monterey, California 93943-5001
831-656-7582
831-656-7599 (fax)

September 30, 2005

Purpose and Rationale

The overarching objective of this program is to assist in the recommendation of the minimum Required Navigation Performance (RNP) value for a VFR helicopter equipped with an IFR GPS. The results of this study combined with the output from another AAR-100's Vertical Flight project entitled "Helicopter SNI helicopter Flight Data" will assist the Federal Aviation Administration flight standards office in determining the minimum RNP value that will be accepted by air traffic office in developing procedures for VFR SNI routes. By correlating human performance data in the simulator to already collected flight data, we will be able to further experiment with new flight patterns towards a decreased minimum RNP value. The purpose of our project is to build and validate the simulation system for further experimentation.

Methodology

A critical element of our study involves a model of pilot performance as a factor of pilotage cues (e.g. landmarks) and radio communications (e.g. GPS receivers). We need to know if a pilot fixates on landmarks versus GPS output. Do they simply "fly the needle" off of the GPS unit, do they carefully observe visual cues, or is it some mix of both? How does this affect the envelope we can assume they are maintaining, therefore indicating how traffic can be controlled around them? We assume that too much attention to the GPS receiver may adversely affect pilotage performance, but that the reverse may also be sub-optimal. The study conducted in this program investigates in a virtual environment simulation how traffic density, workload, and weather affects the minimum RNP for a qualified VFR helo pilot equipped with an IFR GPS.

Recent Accomplishments

The primary accomplishments for this period involve attempts to integrate a KLN-89B GPS emulation system in the simulation and writing analysis tools.

KLN-89 emulation system integration

Work continues on integrating a GPS emulation system into the simulation. The system (KLN-89 receiver emulation board and KLN-89 panel mounted GPS) was delivered in late July. Progress included modifying the sample software delivered with the system to

work in standalone mode. The sample software delivered with the system was tightly coupled and dependent on a larger simulation package. As this package was not included in the delivery, NPS engineers modified the software to run independently. Additional progress included building the required interconnection wiring. After extensive troubleshooting of the hardware and software systems, Frasca technicians determined there was a hardware fault that could only be fixed at Frasca. The system was shipped in early September. Technicians have thus far been unable to repair the system.

Preliminary simulator data analysis

Preliminary investigation of the simulator data was conducted. The results for a sample subject on the first 5 legs of the flight are shown in Figure 1. The horizontal axis represents elapsed time along each leg in seconds; the vertical axis represents magnitude of deviation in meters. Most of the deviation from prescribed route is associated with turns at the waypoints. This can be attributed in part to the GPS emulation system used in the simulation. The GPS used in the aircraft used fly-by waypoints and featured a turn anticipation function. The GPS emulation used in the simulation was based on fly-over waypoint navigation and did not include a turn anticipation feature. Further data analysis and comparison with in flight data will be completed after the GPS emulation system is integrated into the simulator.

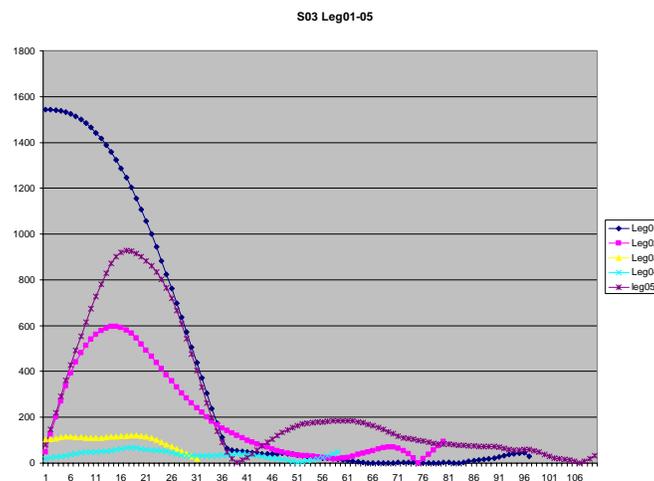


Figure 1. Simulator subject three route deviation for legs 1-5

Flight path and pilot's view visualization tool

To assist with data analysis a post flight/simulation tool was developed. The tool was designed to allow evaluators to visualize navigation performance and the pilot's out the window view. A screen capture from this program is shown in Figure 2. The program inputs a subject's recorded flight path data from either the aircraft or simulator events data files. The tool replays the flight depicting the pilot's out the window (OTW) view in the upper portion of the monitor and its progress along track in the lower portion. Aircraft progress along its flight path is controlled via keyboard entry. The lower window is manipulated with the mouse using a trackball (world in hand) metaphor. Waypoints and the aircraft's current position are depicted as opaque red spheres.

Waypoints are labeled; the sphere associated with the aircraft position is not labeled.
 Table 1 summarizes the functionality and symbology of this tool.

Upper third of screen	Three panels corresponding to left, center and right cockpit out the window views. Keyboard entry allows user to pause/resume flight.
Lower third of screen	Exocentric view of the aircraft's flight path and current position and orientation relative to waypoints.
Symbology	Opaque white band represents the aircraft's track. Opaque red sphere with text label represents the waypoint. Opaque red sphere without label represents the current aircraft position.

Table 1. Summary of flight visualization tool functionality

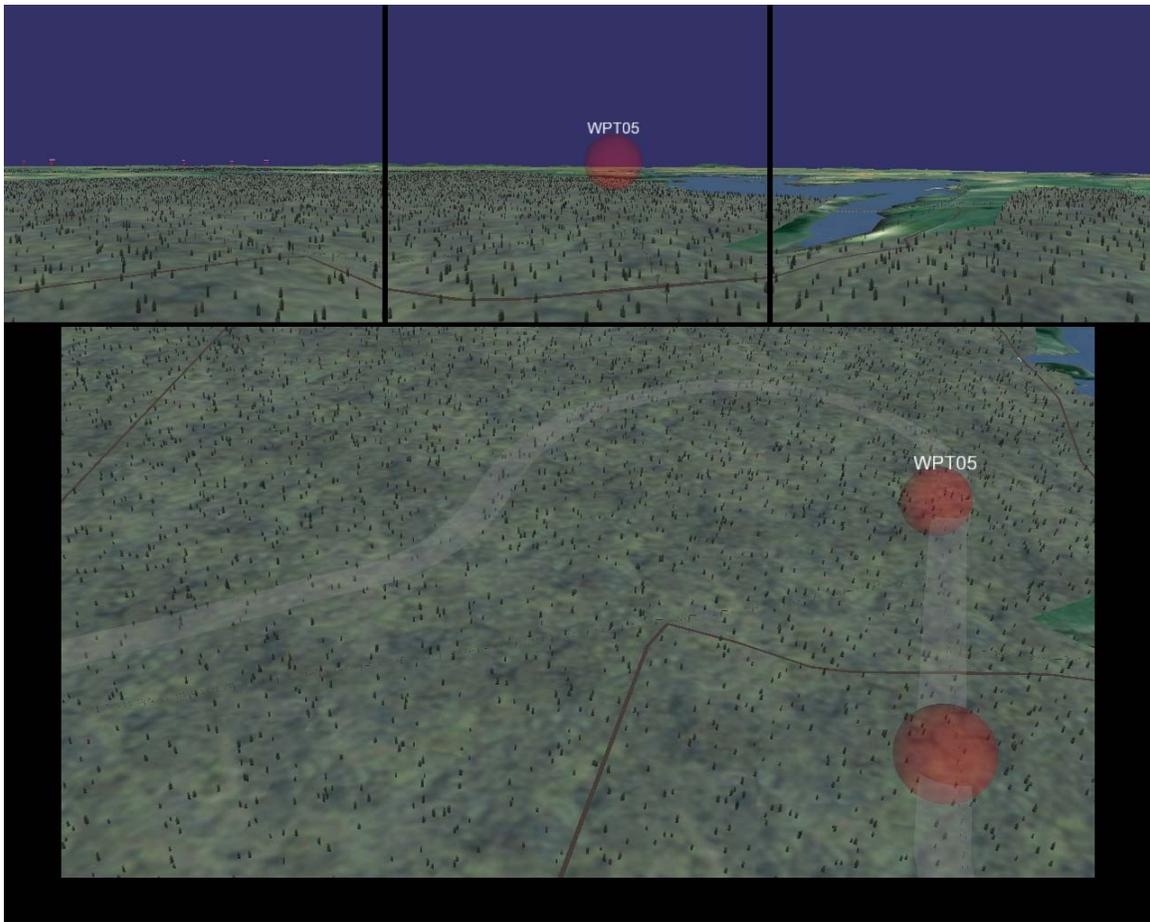


Figure 2. Flight Path, Pilot's View Visualization Tool

The visualization tool may be useful for providing insight into pilot performance. For example in Figure 2 subject three flew well past waypoint five. The visual cues associated with the next segment of flight (a set of power lines) as depicted in the OTW view are not very prominent. This compares to Figure 3 where subject three approaches waypoint eight. Here the feature associated with the next leg of the flight is clearly visible well in advance.

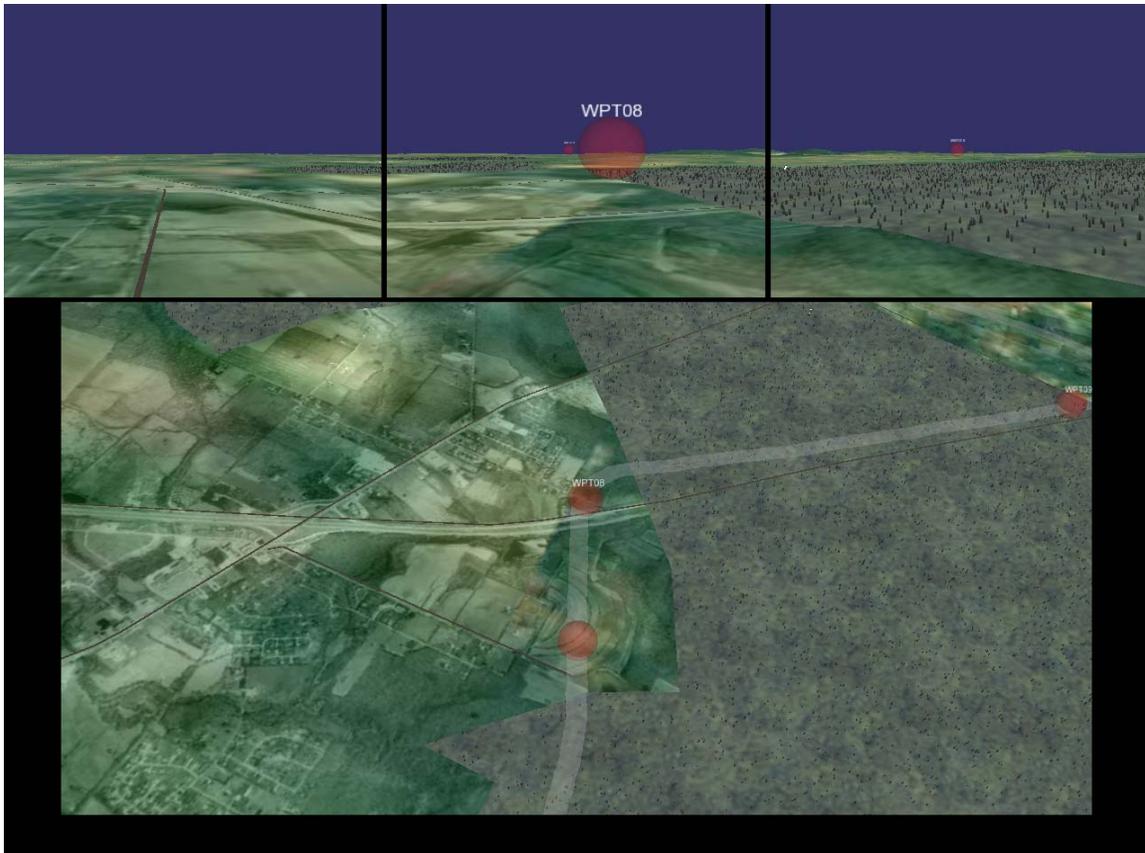


Figure 3. Flight Path & Pilot's View - Subject three approaching waypoint 08

Potential additional features to add to this visualization tool would be a simulation of the the GPS panel. This would provide a depiction of the navigation data available to the pilot throughout the flight. These displays could then be integrated with pilot scan data. The pilot's view through the flight could then be superimposed on the pilots combined out the window and cockpit views.

UAV See and Avoid Systems: Modeling Human Visual Detection and Identification

Andrew B. Watson
NASA Ames Research Center, Moffett Field, California

The FAA seeks to characterize the ability of UAV viewing systems to support target detection and identification. Existing system evaluation methods require expensive and time consuming subjective experiments. We hope to replace those experiments with the Spatial Standard Observer, a simple model of human detection and discrimination. This report describes progress on two elements of this project: simulation of an existing subjective data set using the Spatial Standard Observer (SSO), and development of a web-based application for demonstrating SSO-based visibility calculations. Preliminary results indicate the utility of both elements.

Introduction

The FAA seeks to compile and review the characteristics and performance of existing optical/digital viewing systems that could be used to enhance the human UAV operator's ability to see-and-avoid potential conflicts with other manned and unmanned aircraft. The systems will be characterized by their performance characteristics: field-of-view, field-of-regard, modulation transfer function, focal point, and lens quality, as well as bandwidth and compression. This comparison will be used to determine the ability of these systems to allow detection of static images of differing sizes, at a range of distances in, variety of visibility conditions, i.e., sense-and-avoid.

In this context there is a need to supplement the Army's target acquisition model with a human vision model to predict observers' probability of detection and recognition of aircraft and other targets. In the current Army target acquisition model, these tasks are associated with particular values of N50 for particular image sets and classes, which are obtained by expensive and time consuming subjective experiment. We propose to create and evaluate a tool for computing N50 from a given image set and given classifications, thus obviating the need for subjective measurements. The predicted N50s would be entered in the Army's target acquisition performance model, Night Vision Thermal Imaging System Performance Model (NVTherm), to determine the effects of camera field-of-view, camera field-of-regard, camera modulation transfer function, opposing aircraft size, contrast, distance,

and atmospheric conditions on observers' detection and recognition of an aircraft[1].

We have developed a model called the Spatial Standard Observer (SSO) that allows predictions of visual detection and discrimination of foveal spatial targets (Watson & Ahumada, 2004). The goal of this project was to assess the feasibility of using the SSO to compute N50 values for target image sets.

The first effort in this project has been to simulate the results of a recent psychophysical experiment that estimated N50 for a set of military vehicles[2]. A second concurrent effort has been the development of a prototype tool for calculation of the visibility of manned or unmanned aircraft under specified viewing conditions.

Target Identification Model

Here we describe the development and evaluation of a model to predict image and object identification. We begin with a description of the experiment whose data will be modeled.

Psychophysical Experiment

The experiment has been more extensively described in another report[2]. Here we provide a brief summary. The experiment consisted of two parts, using visible and infrared imagery respectively.

In each part of the experiment, the source images consisted of 144 digital images, of 12 "objects" in 12 "aspects." An illustration of two of the objects and three of the aspects are shown for the visible and infrared imagery in Figure 1. Each

object is a particular military vehicle, and each aspect is a view of that vehicle. The twelve aspects are approximately the same from vehicle to vehicle. Of the twelve aspects, eight are views from an elevation of seven degrees, while the remaining four are from 0 degrees.

These source images were blurred with Gaussian kernels of 6 possible scales,

$$G(\mathbf{x}) = \text{Exp}\left(\frac{-\pi|\mathbf{x}|^2}{\text{scale}^2}\right) \quad (1)$$

The scales ranged from 5 to 30 pixels in steps of 5. This yields a total of $6 \times 144 = 864$ images for each image set (visible or infrared). The six

levels of blur are illustrated in Figure 2.

Identification experiments using trained human observers were run separately on each level of blur. Each observer viewed a subset of 144 images of one type (visible or infrared), consisting of 2 aspects for all 12 objects in all 6 blurs. The two aspects were chosen in a quasi-random fashion. The observers were previously trained on identification of these vehicles, using different images. On each trial, the observer attempted to identify the object. The percent correct was recorded. The results are shown in Figure 3.



Figure 1. Example images. Two objects (rows) and four aspects (columns) are shown for both the visible and infrared image sets. The last aspect shows an example of the 0 degree elevation.



Figure 2. Examples of the six levels of blur applied to one image of each type (visible and infrared).

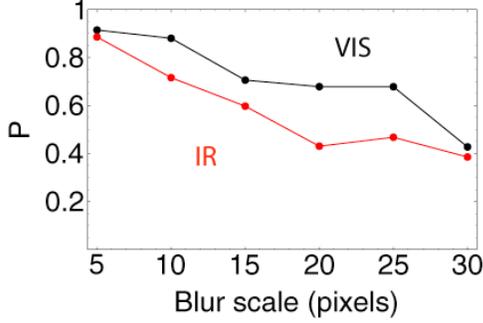


Figure 3. Percent correct identification as a function of blur scale for visible and infrared targets.

Model

The first model we have considered is a simple image classification machine operating on the basis of a normalized correlation matching rule. This model computes a set of N discriminant functions, where N is the number of possible images (in this case, $N = 144$). One discriminant corresponds to each candidate image, and the model selects the image with the largest discriminant.

The matching is assumed to occur in a “neural image” space, which is reached by transforming the image. The transformation consists of a conversion to contrast and filtering by a contrast sensitivity filter (CSF). The CSF is derived from our Spatial Standard Observer (SSO), a simple model of foveal contrast detection[3].

The templates consist of the transformed images. If the presented transformed image is written s (for sample), then the discriminant for image i is given by

$$d_i(s) = s \mathbf{g}_i \quad (2)$$

where t_i is the normalized template. It is not necessary to divide by the norm of s , since it is the same for all discriminants.

Each transformed image can be expressed as a product of its normalized form and its energy

$$g_k = e_k t_k \quad (3)$$

Thus if image k is presented,

$$s = e_k t_k + n \quad (4)$$

where n is a neural noise image (noise in the neural image space). Then

$$\begin{aligned} d_i(s) &= (e_k t_k + n) \mathbf{g}_i \\ &= e_k t_k \mathbf{g}_i + n \mathbf{g}_i \end{aligned} \quad (5)$$

We can divide through by e_k without changing the ranking of the discriminants,

$$\begin{aligned} d_i(s) &= t_k \mathbf{g}_i + \frac{n \mathbf{g}_i}{e_k} \\ &= \rho_{i,k} + \frac{n \mathbf{g}_i}{e_k} \end{aligned} \quad (6)$$

where $\rho_{i,k}$ is the correlation (dot product) between each pair of neural images.

If the noise is white and normally distributed with standard deviation σ , then the second term in this expression will be a normally distributed random variable with standard deviation σ/e_k . So finally, each discriminant will be a normal random variable distributed as

$$d_i(s) = \text{Normal} \left(\rho_{i,k}, \frac{\sigma}{e_k} \right) \quad (7)$$

To simulate performance of this model, we simply pick a noise σ , and generate N discriminant values for a number of trials T for each of N sample images. On each trial, the image selected is the largest discriminant, and from these results we can compute percent correct (we can also generate confusion matrices). We compute both percent correct image identification and correct object identification. The performance of the model is controlled by a single parameter: σ , the standard deviation of the “neural noise” added to the sample neural image. In Figure 4, we plot the percent correct for image identification and object identification for images blurred by 30 pixels.

As expected, increasing noise reduces performance. The red and green lines in the figure show the asymptotic guessing performance expected given the numbers of images and objects, and the larger values of noise reach these asymptotes.

Another question of interest is whether the image and object identification performance can

be related by a simple guessing model: is the object identification performance what would be expected by assuming that if the model does not pick the correct image, that it then guesses among the other images. In that case the percent correct object identification (P_o) can be computed from the percent correct image identification (P_I) as

$$P_o = P_I + \left(1 - P_I\right) \frac{N - 1}{N^2 - 1}. \quad (8)$$

This prediction is shown by the gray curve in Figure 4. Clearly, in this example, the object identification is better than would be expected from this prediction. We call this the "object advantage" (OA). The OA is negligible at 5 pixels blur, but increases to a max of about 0.13 at 30 pixels. Without an aperture (see below), it is about the same for VIS and IR. With an aperture, it is smaller for IR than for VIS. Possible sources for the OA are: background (without aperture), object color (for visible), and overall object size. We will return to this point later.

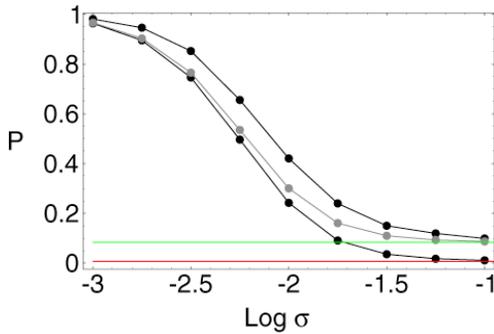


Figure 4. Percent correct image (lower black curve) and object (upper black curve) identification for various levels of the noise standard deviation. These results are for visible targets at blur scale = 30 pixels. Green and red lines indicate predicted guessing performance. The gray curve is object identification predicted from image identification using a guessing model (see text).

Object Identification vs Blur Scale

The results for image identification can also be plotted as a function of blur scale, as shown in Figure 5. The value plotted is percent correct object identification (as in the upper curve in Figure 4), and each curve is for a different noise sigma.

The figure also includes (blue and red curves) the data from the human observers. No attempt has been made at this point to find the best fitting value of noise σ , but it is clear that a value of around -2.25 yields a rough approximation to the human data for visible images, and -2 for infrared images.

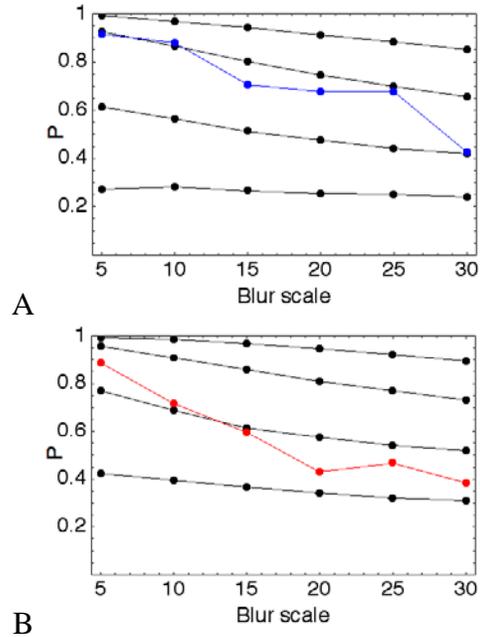


Figure 5. Simulated percent correct object identification as a function of blur scale for several different values of neural noise ($\text{Log } \sigma = -2.5, -2.25, -2., -1.75$). The blue and red curves are the human data. A) visible, B) infrared.

Removing the Background

As noted above, object identification performance is better than expected from the guessing model, which indicates that on average different aspects of one object are more similar (as images) than are aspects of another image. This could be due in part to the object background, which is nearly constant from aspect to aspect. To test this we have computed results for images with the background removed. Aperture images defining the object area were provided by the U.S. Army Night Vision and Electronic Sensors Directorate. The apertured image was constructed as $\text{image} * \text{aperture} + 2048 * (1 - \text{aperture})$. An example of the construction of one apertured image is shown in Figure 6.

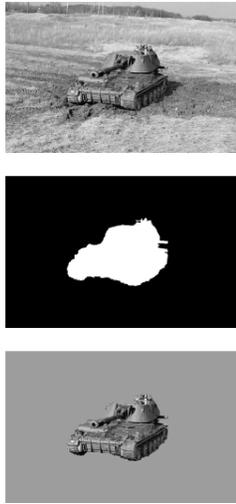


Figure 6. Construction of an apertured image. A) Original image, B) aperture, C) apertured image.

The model results obtained using the apertured images are shown in Figure 7. Overall, performance is somewhat better than for the original images. The visible image performance for $-\text{Log } \sigma = -2.25$ is now closer to the data, while the infrared data lie between $\text{Log } \sigma = -2.5$ and -2.25 .

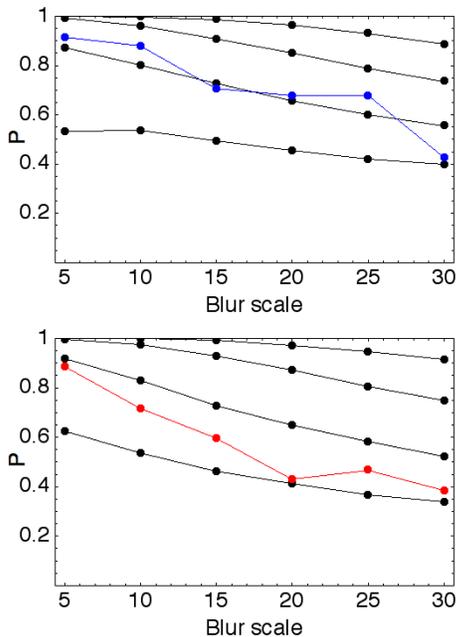


Figure 7. Object identification performance vs blur scale for apertured images. Details as in Figure 5.

Visible vs Infrared

One purpose of the original psychophysical experiment was to determine the relation between N50 for visible and infrared images of similar objects. If the N50s were the same, that would allow the same metric to be used regardless of the image type. However, in that experiment the estimated N50s differed by about 50% (7.5 visible, 11.5 infrared)[2].

Figure 8 compares model results for visible and infrared. A short summary is that performance is somewhat better for infrared than for visible, but that this advantage largely vanishes with apertured images. Recall that human performance is slightly lower for infrared, so this constitutes a small discrepancy between model and data.

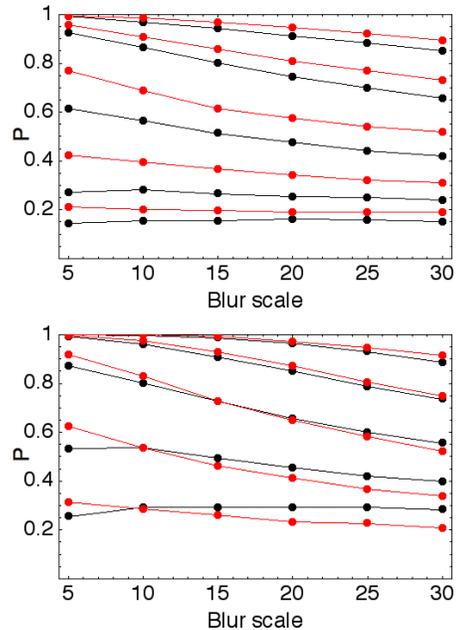


Figure 8. Object identification performance vs blur scale for visible (black) and infrared (red) images. A) Original, B) apertured. Other details as in Figure 5.

Summary

A very simple identification model incorporating the Spatial Standard Observer can generate performance similar to human data for both visible and infrared imagery. Some discrepancies remain, notably the slightly steeper decline with blur, and the poorer performance with infrared

imagery, found in the human results. We hope to investigate these matters further in the second stage of this project.

Future work on this part of the project will include alternative SSO-based models, as well as other human data sets[4]. We hope to understand better the reasons for infrared vs visible performance. We also want to work with aircraft rather than tank images.

Visibility Calculator

In a second part of this project, we have begun development of a prototype application to predict visibility of aircraft targets as they might be seen from a UAV. Conversely, the tool could be used to predict visibility of the UAV from another aircraft. A screen shot of the prototype application is shown below.

The tool allows the user to select an aircraft, as well as various viewing parameters. The tool then computes the visibility of the aircraft, expressed in units of JND. The tool is currently online and operational at the URL shown in the figure.

The tool operates by computing a rendered image from a selected 3D model. The rendered image is then processed using the current version of the Spatial Standard Observer (SSO). The tool is implemented using webMathematica, an extension of the Mathematica language[5]. The current version of the prototype is only a proof of concept, and must be augmented by realistic optical and atmospheric effects, and must be calibrated in both geometric and photometric aspects. We plan to accomplish these augmentations in the second phase of this project.

Acknowledgments

I thank Albert J. Ahumada, Jr. and Jeffrey B. Mulligan for helpful discussions, and thank Ron Driggers and Eddie Jacobs of the Army Research Lab for providing the target images and background information on the psychophysical experiment.

References

1. Vollmerhausen, R.H. and E. Jacobs, *The Targeting Task Performance (TTP) Metric A New Model for Predicting Target Acquisition Performance*, Modeling and Simulation Division Night Vision and Electronic Sensors Directorate U.S. Army CERDEC Fort Belvoir, VA 22060.
2. Driggers, R.G., et al., *Fifty-percent probability of identification (N_{50}) comparison for targets in the visible and infrared spectra*, US Army Night Vision and Electronic Sensors Division (NVESD).
3. Watson, A.B. and A.J. Ahumada, Jr., *A standard model for foveal detection of spatial contrast*. Journal of Vision, in press.
4. Moyer, S., et al., *Cycle Criterion for Fifty-Percent Probability of Identification for Small Handheld Objects*. 2005, US Army CE-COM, RDEC Night Vision and Electronic Sensors Directorate Ft. Belvoir, VA.
5. Wolfram, S., *The Mathematica Book*. 5th ed. 2003, Champaign, IL: Wolfram Media.

See and Avoid

http://eveleth.arc.nasa.gov:8080/webMathematica/abw/saa.jsp

NASA "See and Avoid" Visibility Calculator

Aircraft:

Background (0-1):

Ambient lighting:

Light source distance: Color:

Distance:

Viewpoint:



View of the selected aircraft under the specified viewing conditions. Click and drag on the image to rotate, shift-click-drag vertically to zoom.



View with addition of cloud background.



View from specified distance.

Visibility = 16.5641 JND.

Uncalibrated Prototype.

Date = {2005, 9, 16, 17, 5, 32.354173}

Mathematica Version = 5.2 for Mac OS X (64 bit) (June 20, 2005)

Figure 9. Screen shot of web-based visibility tool.

VIDEO PROCESSING METHODS FOR IN-FLIGHT GAZE ANALYSIS

Jeffrey B. Mulligan
NASA Ames Research Center

In-flight gaze analysis is a tool for assessing the impact of new cockpit technologies on pilots' allocation of attentional resources. In particular, gaze tracking measures allow us to determine whether external scanning is sufficient to insure the pilot's ability to see-and-avoid traffic under VFR conditions. Commercial gaze-tracking solutions, however, do not currently provide adequate performance in the presence of high levels of ambient illumination, as encountered in clear sunny weather. This report describes novel methods developed for the analysis of data collected in a series of helicopter flight tests conducted in October, 2003.

INTRODUCTION

Pilots flying under visual flight rules (VFR) are obligated to continuously monitor the surrounding airspace for other traffic, and maneuver as necessary to eliminate conflicts ("see-and-avoid"). Thus the introduction of any new device into the cockpit raises the question of how the use of the device may impact the pilot's allocation of visual and attentional resources. Our project specifically focusses on the use of global positioning system (GPS) receivers as navigational aids. We wish to determine both how access of the information provided by the display affects performance in a precision navigation task, and how it impacts other important functions such as see-and-avoid. The use navigational aids of this sort is of particular importance for helicopter operations such as medical evacuation, in which the pilot has to fly an unfamiliar route, possibly in close proximity to obstacles and other traffic.

To this end, a series of flight tests were conducted in October, 2003, in which four video streams were recorded. Two head-mounted cameras provided images of the pilot's eye, and the forward-looking view from the pilot's perspective, while two additional fixed cameras provided a frontal view of the pilot's head and shoulders, and an over-the-shoulder view which included the control stick. A complete description of the data collection procedures has been reported previously [1].

Our initial approach to extraction of gaze estimates from the video data was to apply techniques commonly applied to similar images obtained in the laboratory under controlled

illumination conditions [2]. Unfortunately, these techniques proved unsatisfactory for the conditions encountered during the flight tests. Straightforward search for key features such as the illuminator reflexes ("glints") and the pupil boundary (inner iris margin) resulted in gaze estimates for approximately 70% of the frames in the night recordings, and less than 40% of the frames of the day recordings. The primary factor contributing to the poor performance with the day recordings was the high level of ambient illumination (sunlight) which swamped the controlled illumination provided by the goggle-mounted light-emitting diodes. Additionally, the high light levels caused most of the subject pilots to maintain their eyelids in a relatively closed position, often hiding the features normally used for gaze estimation.

In order to obtain precise gaze estimates for all of the images, we therefore embarked upon a program to develop a set of new methods specifically tailored to address these problems. Our approach consisted of the following elements: 1) development of interactive tools for hand-labelling of selected images; 2) development of a geometrical model of the eye, allowing accurate gaze estimation from a minimal set of features; 3) development of a clustering procedure for selecting minimal sets of exemplar images for hand labelling, which span the space of possible appearances; 4) development of interactive tools for registration and feature-labelling of images from the head mounted scene camera, necessary for transforming head-relative gaze estimates (obtained from the eye images) to an external world-referenced gaze target. In the following

sections, we describe each of these elements in more detail.

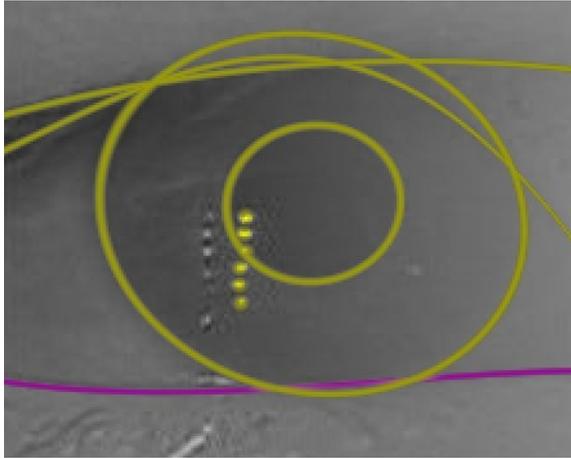


Figure 1: Typical eye image (from night flight) showing superimposed labels of eyelid and iris features.

EYE AND LID LABELLING TOOL

The eye and lid labelling tool allows an operator to indicate the positions of the features of interest with a series of mouse clicks within a window displaying an enlarged image. The set of possible features consists of: 1) three fourth-order curves describing the lower eyelid margin, the upper eyelid margin, and the skin fold above the upper eyelid; 2) two ellipses describing the inner and outer margins of the iris, referred to as the pupil and limbus, respectively; 3) six point locations describing the positions of the reflections of the LED illuminators. Additionally, check-boxes are provided allowing the operator to indicate the presence or absence of each feature in each image to be labelled. Figure 1 shows a typical image in which all the features are visible, along with the corresponding labels.

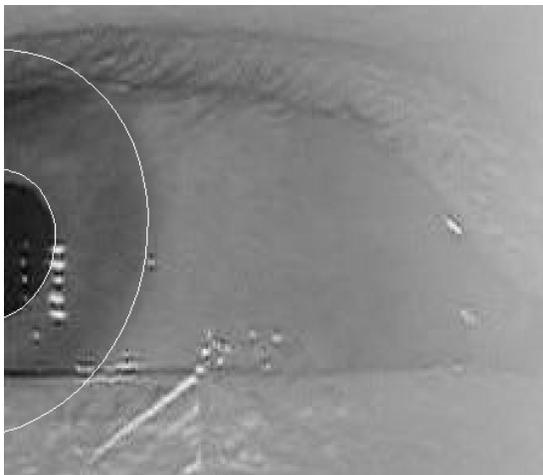


Figure 2: Eye image with superimposed labelling showing pupil/limbus model.

GEOMETRICAL EYE MODEL

The labelling procedure described in the previous section allows the pupil and limbus to be described by ellipses which are completely independent. But because these features are part of a rigid mechanical system (the eye), they move together, and thus their projected shapes in the image are not free to vary independently, but are strongly constrained. These constraints may be exploited to obtain accurate estimates of gaze even when only a small portion of the pupil is visible in the image (as in figure 2).

We have implemented a model introduced by Ohno [3], in which the effects on the pupil image by refraction at the cornea are approximated by a change in apparent depth and size. The model has 3 structural parameters, which should be the same for all images obtained from a given subject: the limbus radius, the distance of the plane of the iris from the eye's center of rotation, and the difference in apparent depth between the pupil and the limbus. Two additional parameters are constant within a set of images obtained with a fixed position of the goggle: these are the position in the image of the center of the pupil and limbus when the eye is pointed directly at the camera, and the pupil and limbus appear as concentric circles. The corresponding viewing direction forms the origin of our gaze coordinate system.

Three additional parameters must be determined for each frame: the gaze angles, expressed as slant and tilt relative to the eye-camera axis, and the pupil radius (which varies slowly within limits). As slant increases, the pupil and limbus change in appearance from circles to ellipses; the major axis of the ellipse having a length equal to twice the relevant radius, while the minor axis is diminished by a factor equal to the cosine of the slant. If the pupil depth

difference parameter is zero, then the ellipses will be concentric; conversely, the depth difference parameter can be adjusted to account for non-concentric appearance at large gaze angles.

Several passes through the data are required to determine the fixed parameters: first we must determine the center coordinates. If the model is accurate, then all of the ellipse minor axes should intersect at the center point. In practice, the ellipses produced by the initial labelling will not have coincident minor axes, so we obtain a least-squares solution using the singular value decomposition on the matrix of line equations. Once the correct center has been found, then the shape of the limbus in the frames with large gaze deviations determines the distance of the limbus plane from the center of rotation. Finally, the offset of the pupil plane is readjusted in each frame. In each case, after labelling the individual frames, the mean is computed across frames, and this value is held fixed during subsequent iterations. Once the structural parameters have been determined, the variable parameters (gaze angles and pupil radius) can be set quickly and easily.

IMAGE CLUSTERING

While the hand-labelling procedures described in the previous sections require only a minute or so per frame, when the number of frames is large it is impractical to hand label them all. For example, the Tullahoma flight test data set consists of 15 recordings of approximately 100,000 frames each. Fortunately, many of the frames are roughly similar; because typical gaze behaviors consist of fixational eye movements, we often encounter runs of 10 or more similar frames corresponding to a fixation. Furthermore, because gaze repetitively returns to certain targets such as the cockpit instruments, we find large sets of similar frames in the complete recordings.

The purpose of the image clustering procedure is to form an efficient hierarchical representation of this structure.

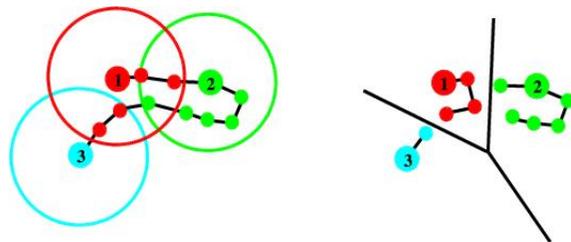


Figure 3: Two-dimensional cartoon illustrating selection of catalog exemplars and nearest neighbors. Numbered disks represent exemplar images, a new exemplar is added to the catalog when the distance of a new image from existing exemplars exceeds a threshold, indicated by the large circles.

The procedure we have adopted combines elements of vector quantization [4] and nearest-neighbor classifiers [5]. Here we present a brief overview of the procedure; a more thorough treatment is provided elsewhere [6]. We assume the existence of a metric which provides us with a measure of “distance” between two images. (We use a metric based on correlation, but the following discussion does not depend upon the choice of metric.) We treat the images as points in a high-dimensional space; the number of dimensions is potentially as large as the number of pixels, but for the restricted class of images that we are dealing with the images all lie within a manifold whose dimension is considerably lower. In figure 3, we represent the images as points in a two-dimensional plane for illustration purposes only.

We begin by choosing a threshold distance. Our goal is to come up with a minimal set of exemplar images, chosen from the input sequence, such that each exemplar differs from every other exemplar by at least the threshold distance, but every other non-exemplar image is within the threshold distance of the nearest exemplar. The catalog of exemplars is formed as follows: the first image in the sequence is

the first catalog entry. As we proceed sequentially through the sequence, each image in the sequence is then tested against the exemplar associated with the previous frame. If the distance is below the threshold, then we proceed to the next frame. Otherwise, we test the image against the remaining catalog entries, stopping when we find one whose distance from the input is less than the threshold. If no catalog entry is found within the required distance of the input image, then the input image is added to the catalog. After the catalog has been generated, a second pass over the data is performed in which each image is associated with its nearest neighbor in the catalog. This process is illustrated in figure 3. Rather than process the entire sequence with a small threshold, we begin with a large threshold resulting in a small number of exemplars, and then apply the process recursively to the resulting neighborhoods, resulting in a tree in which the exemplars at each levels form the nodes. As we descend the tree, the images in each neighborhood become more and more similar; at some point we expect that this similarity will be high enough that an automatic labelling procedure, initialized with the values of a hand-labelled exemplar, will be able to successfully label the remaining images.

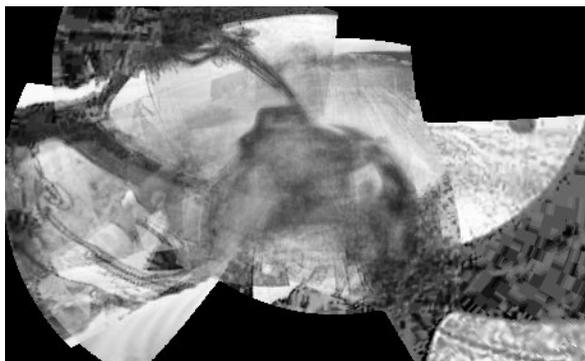


Figure 4: Cockpit mosaic image created by merging hand-aligned exemplar images.

Analysis of the eye images as described above tells us the direction of gaze relative to the head; similarly, the head-mounted scene camera provides us with a head-relative view of the world; each pixel in the scene camera image corresponds to unique direction of head-relative gaze. Thus, once we have registered the scene camera image to a model of the world, we can relate the gaze computed from the eye image to an external target specified in world-coordinates.

Initially, we make the assumption that translations of the scene camera are small compared to the distance to the objects being imaged, so that we can ignore the effects of parallax, and model the appearance of the cockpit by mosaicking images from the scene camera on a sphere. We have developed a tool allowing an operator to manually register an image to another image or the complete mosaic by manipulating sliders controlling the three rotation angles (pan, tilt, and roll). This is accomplished by first computing the angles associated with each pixel in the scene camera image (which depends only on the focal length). These angles are then transformed according to the operator-selected parameters. An entire hemisphere of viewing directions is mapped into an image for viewing using stereographic projection (see figure 4).

While it is possible to obtain a reasonable looking mosaic in this way, individual features are often misaligned. This can be for two reasons: first, our assumption of zero parallax is clearly false; in addition, the focal length of the camera is initially uncalibrated. Both of these issues are ones which we ultimately hope to deal with in the correct manner, but in order to do so we need to have the coordinates of individually-labelled features. Thus our labelling tool also incorporates a feature editor. To add a new feature, the operator first clicks on its location in the mosaic image. The tool then automatically generates the list of frames which should contain that feature, based on the angles used to register each frame to the mosaic. These images are presented to the user in a second window, where (s)he indicates the precise location with another mouse click.

SCENE LABELLING TOOL

SUMMARY AND CONCLUSIONS

We have described a number of new tools developed to aid the analysis of in-flight gaze recordings. Currently, a few thousand images from the Tullahoma flight tests have been hand-labelled, which allow us to directly estimate gaze, but with a low precision. Because of the fact that the GPS receiver was mounted at the top of the instrument cluster in the test vehicle (i.e., at the boundary of the windscreen), we need a high degree of precision to discriminate fixations on the receiver from out-the-window scanning. Thus our next step will be to develop procedures to use the hand-labelled data to automatically label the remaining images.

REFERENCES

1. Mulligan, J. B. (2005). Pilot Behavior and Course Deviations during Precision Flight, in Rogowitz, B. E., Pappas, T. N., and Daly, S. J., (eds.), *Human Vision and Electronic Imaging X*, Proc SPIE vol 5666, pp. 363-373.
2. Mulligan, J. B., (1997). Image Processing for Improved Eye-tracking Accuracy. *Behavioral Research Methods, Instrumentation and Computers*, vol. 29, pp. 54-65.
3. Ohno, T., Mukawa, N., and Yoshikawa, A. (2002). FreeGaze: a Gaze Tracking System for Everyday Gaze Interaction, in Duchowski, A. T., Vertegaal, R., and Senders, J. W. (eds.), *Proc. ACM Symposium on Eye Tracking Research and Applications (ETRA)*, pp. 125-132.
4. Gersho, A. and Gray, R. M. (1992). *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Norwell MA.
5. Cover, T. M. and Hart, P. E. (1967). Nearest Neighbor Pattern Classification, *IEEE Transactions on Information Theory*, vol. IT-13, pp. 21-27.
6. Mulligan, J. B. (in press). A Tree-structured Model of Visual Appearance Applied to Gaze Tracking. Proc. 2005 IEEE International Symposium on Visual Computing.